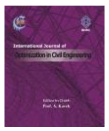
INTERNATIONAL JOURNAL OF OPTIMIZATION IN CIVIL ENGINEERING
Int. J. Optim. Civil Eng., 2025; 15(1):39-56



A COHERENCY METRIC TO COMPARE OPTIMALLY CLUSTERED SEISMIC DATA

M. Shahrouzi[*, †], M. Rashidi-Moghaddam
Civil Engineering Department, Faculty of Engineering, Kharazmi University, Tehran, Iran

[DOI: 10.22068/ijoce.2025.15.1.621]

ABSTRACT

Clustering is a well-known solution to deal with complex database features as an unsupervised machine learning technique. One of its practical applications is the selection of non-similar earthquakes for consequent analysis of structural models. In the present work, appropriate clustering of seismic data is searched via optimization. Silhouette value is penalized and used to define the performance objective. A stochastic search algorithm is combined with a greedy search to solve the problem for distinct sets of near–field and far-field ground motion records. The concept of coherency is borrowed from optics to propose a coherency metric for earthquake signals before and after being filtered by structural models. It is then evaluated for various cases of structural response-to-record and response-to-response comparisons. According to the results the proposed coherency detection procedure performs well; confirmed by distinguished structural response spectra between different clusters.

Keywords: Coherency Index, clustering, soft computing, optimization, nonlinear dynamic analysis.

Received: 23 December 2024; Accepted: 26 January 2025

1. INTRODUCTION

As an interdisciplinary field, data mining brings together the techniques from machine learning, pattern recognition, statistics, databases and visualization to address the issue of information extraction from large data bases. Some major tasks that data mining is usually called upon to accomplish, are: description, estimation, prediction, classification, clustering and association [1].

Clustering is the task of partitioning the database into some groups (clusters) such that the entities are very similar within every cluster, but as dissimilar as possible to those of the others.

[*]Corresponding author: Civil Engineering Department, Faculty of Engineering, Kharazmi University, Tehran, Iran
[†]E-mail address: shahruzi@khu.ac.ir (M. Shahrouzi)

Depending on the data and desired cluster characteristics, there are different types of clustering paradigms such as representative-based, hierarchical, density-based, graph-based, and spectral clustering [2]. Representative-based clustering aims at finding a set of K representatives that best characterize a dataset [3]. It includes the well-known K-means as a greedy algorithm that minimizes the error on entity distances from their respective cluster centers. A good clustering leads to the best value of a desired measure; therefore, it can be considered as an optimization task [4].

In the seismic design that employs time-history analysis, it is crucial to obtain an appropriate set of ground motion records for an accurate estimation of the dynamic structural responses under the given hazard level at the construction site. Availability of online digital databases of earthquakes has increased accessibility to real-world ground motions; however, depending on the recording station, earthquake magnitude, faulting type, soil condition, strong pulse duration and source-to-site distance; the ground motion records can have very different spectral characteristics. In order to take the seismic hazard of the site into account, one has to obtain ground motions that best comply with a specific hazard scenario for that region as specified by the design code.

A number of investigators have addressed optimal selection or scaling of ground motion records so that their mean spectrum best matches a given design target within a period range of interest [5–7]. In practice, the aforementioned attempts do not offer more than a record set. The matter is concerned here via optimal clustering of the records; resulting in more than one option in selecting earthquakes from each of the clusters.

The present work utilizes a hybrid optimizer framework for clustering on two distinct sets of data. First, the earthquake records with geotechnical and seismic attributes and second the structural responses generated via nonlinear analyses under such records. Furthermore, a coherency measure is offered and applied for the optimal clusters in distinct cases of structural response-to-record and response-to-response comparisons. The proposed method is evaluated on a number of moment frame examples; followed by the discussion of the results.

[DOI: 10.22068/ijoce.2025.15.1.621]

2. PROBLEM DEFINITION

One of the most popular indices for clustering desirability is the *silhouette* value. It is a measure of how similar an object is to its own cluster in comparison with the other clusters. The silhouette value ranges from -1 to 1, where higher values show better matching of the entity to its own cluster rather and -1 counts vice versa. It is defined for any i^{th} entity by Eq. (1).

$$s(i)=\begin{cases}1-\dfrac{a(i)}{b(i)} & \text{if } a(i)<b(i)\\ \dfrac{b(i)}{a(i)}-1 & \text{if } a(i)>b(i)\\ 0 & \text{otherwise}\end{cases} \tag{1}$$

For each datum i, $a\ (i)$, is the average dissimilarity of i with all other data within the same cluster. The smaller value of $a\ (i)$, the better the assignment is. Furthermore, $b\ (i)$ is the lowest average dissimilarity of i to any other cluster, of which the i^{th} entity is not a member.

In order to seek the best clustering; a fitness function is defined based on sum of mean silhouette values over all clusters of the database. The optimization problem is formulated as:

$$Max\ Fitness(x_1,...,x_{N_e}) = \frac{\sum_{j=1}^{N_e}\sum_{i=1}^{N_i} s(i)}{N_e} \tag{2}$$

in which, N_i is the number of objects in the i^{th} cluster. Total number of entities; i.e. number of rows in the data matrix is denoted by Ne. An integer number between 1 and K can be assigned to each component of the design vector X; where K stands for the prescribed number of clusters [8].

3. THE CLUSTERING FRAMEWORK

Several clustering solutions have already been introduced by investigators [2,9,10]. They can be categories into hard computing or soft computing approaches. The former can locally reveal a specific solution depending on its starting point while the latter includes some fuzzy or stochastic operators for better global search [11–13]. Consequently, a hybrid solution will be of interest to integrate the aforementioned approaches and get merits of both in the efficiency and the effectiveness.

A hybrid framework is introduced here including a meta-heuristic algorithm in the class of soft computing and a greedy algorithm as a local search engine. It is further employed to derive optimal clusters not only for the seismic-excitation signals but also for consequent structural responses to observe the effect of such a filtering. Details of the algorithms are briefed as follows.

3.1 Deterministic solution for local search

As a deterministic solution; the *K-Means* algorithm, KM, is concerned here. KM assigns entities in the given data to K clusters provided that each cluster is identified by location of its centroid and radius. It performs hard clustering as each point is assigned to only one cluster. In addition, KM can generate convex-shaped clusters.

KM procedure is initiated with K randomly positioned centers. Then, every j^{th} entity: x_j is associated with the nearest centroid μ_{i*}.

$$i^* = \arg\min_{i=1}^{K} \left\| x_j - \mu_j \right\|^2 \tag{3}$$

Such a subroutine is repeated until no entity remains outside the K clusters; provided that no

[DOI: 10.22068/ijoce.2025.15.1.621]

cluster is empty at the same time. The next step is updating the centroid positions by averaging new positions of the entities within each cluster.

$$\mu_i^t = \frac{1}{\|C_i\|} \sum_{x_j \in C_i} x_j \tag{4}$$

The clustering assignment and centroid updating steps are repeated alternately until convergence criterion is met; i.e. no further significant change (more than a given error: ϵ') in center locations is observed. Such an iterative procedure aims to minimize the following SSE score:

$$SSE = \sum_{i=1}^{K} \sum_{x_j \in Cluster_i} \left\| x_j - \mu_i \right\|^2 \tag{5}$$

KM (K, ϵ'):
1. t = 0
2. Randomly initialize K centroids: $\mu_1^t, \mu_2^t, \dots, \mu_k^t \in R^d$
3. **Repeat**
4. t ← t + 1
5. $C_j \leftarrow \emptyset \; for \; all \; j = 1, \dots, k$

// Cluster Assignment Step
6. **For each** $\boldsymbol{x_j} \in \boldsymbol{Data}$ **do**
7. Find the closest centroid C_{i^*} by Eq.3
8. Assign the entity to cluster: $C_{i^*} \leftarrow C_{i^*} \cup \{\boldsymbol{x_j}\}$

// Centroid Update Step
9. **For each** $i = 1 \; to \; k$ **do**
10. Update centroid location by Eq.5
11. **Until** $\sum_{i=1}^{k} \|\boldsymbol{\mu}_i^t - \boldsymbol{\mu}_i^{t-1}\| < \epsilon'$

Figure 1: Pseudo-code for the employed KM subroutine

3.2 Stochastic solution and the hybrid framework

Colliding Bodies Optimization, CBO, is a popular meta-heuristic algorithm with several engineering applications [14–18]. According to CBO analogy, a *body* is a candidate solution or a design vector. Such bodies are subdivided into stationary and moving ones based on their fitness. Moving bodies collide with the stationary ones and their velocities are updated using mechanical laws of momentum conservation and energy absorption. A hybrid *Enhanced Colliding Bodies Optimization* [19] and *K-means* [8] has been introduced that uses an auxiliary memory of the elite bodies during its search and takes benefit of KM via a re-initiation subroutine. Such a hybrid framework (ECBO-KM) is utilized for the current optimal clustering using the following steps.

1) Initiation: Randomly generate a population of *n* colliding bodies within the design variable range. In the present clustering problem, each variable is an integer number between 1 and *K*. The function *rand* generates a random floating-point value between 0 and 1.

[DOI: 10.22068/ijoce.2025.15.1.621]

$$x_{ij} = round\left(1 + rand \times \left(K - 1\right)\right) \tag{6}$$

2) Mass calculation: For every i^{th} colliding body, CB_k calculate its mass after fitness evaluation of the entire population:

$$m_i = \frac{Fitness(X_i)}{\sum_{j=1}^{n} Fitness(X_j)} \tag{7}$$

3) Collision: Half of the population members are denoted as moving CB's. They move toward stationary ones that have zero velocities (the fitter half CB's after sorting the population).

$$V_i = 0 \quad , \quad i = 1, \ldots, \frac{n}{2} \tag{8}$$

In this stage, velocity of every moving CB is determined by:

$$V_i = X_i - X_{i-\frac{n}{2}} \quad , \quad i = \frac{n}{2} + 1, \ldots, n \tag{9}$$

4) Update the coefficient of restitution by

$$\varepsilon = 1 - \frac{Iter - 1}{N_{MaxIter} - 1} \tag{10}$$

In which ε stands for the *coefficient of restitution*, COR. It determines the ratio of relative velocity between CB's after collision to such a relative velocity, before collision.

5) Velocity update: After collision, new velocities of colliding bodies are updated due to the Eq. (11)-(12):

$$V_i^{'} = \frac{\left(m_{i+\frac{n}{2}} + \varepsilon m_{i+\frac{n}{2}}\right) V_{i+\frac{n}{2}}}{m_i + m_{i+\frac{n}{2}}} \quad , \quad i = 1, \ldots, \frac{n}{2} \tag{11}$$

$$V_i^{'} = \frac{\left(m_i - \varepsilon m_{i-\frac{n}{2}}\right) V_i}{m_i + m_{i-\frac{n}{2}}} \quad , \quad i = \frac{n}{2} + 1, \ldots, n \tag{12}$$

6) Update position of CB's after collision:

$$x_i^{new} = \max\left(1, \min\left(K, round\left(x_i + rand.v_i^{'}\right)\right)\right), \quad i = 1, \ldots, n \tag{13}$$

[DOI: 10.22068/ijoce.2025.15.1.621]

7) Update the Colliding Memory: This auxiliary memory is updated by sorting and saving CMS number of best-so-far solutions that are already found, up to the current iteration. They are then replaced by the worst CB's in the current population.
8) Mutation: Each j^{th} design variable is regenerated by Eq.6. as soon as a random number falls below the prescribed threshold P_m.
9) Repeat the steps 2~8 for the iteration numbers:1 to N_{KMS}.
10) Re-initiate the population by calling the KM algorithm as the subroutine. KM takes the current population as input and gives its best resulted clustering as the updated population
11) Repeat the steps 2~8 for the iteration numbers: $N_{KMS}+1$ to $N_{maxIter}$.

Note that according to the problem formulation, each component of the design vector in Eq.13 should be rounded to a cluster number between 1 and K.

4. NUMERICAL SIMULATION

4.1 Optimal clustering of earthquakes due to ground motion characteristics

Entities of the data matrix for clustering are distinguished by differences in their attributes. The number of columns in the data matrix depends on the considered attributes while its rows are limited to Ne; i.e. the number of available earthquake records. Cardinality of the entire search space is dominated by various ways of assigning entities into the specified number of clusters; K.

There are several ground shaking characteristics to reflect source, path, and site effects. Some important attributes are selected here including: earthquake magnitude, mean shear wave velocity in the uppermost 30 meters depth of the soil, epicenter-to-site distance, *Peak Ground Displacement*, PGD and *Peak Ground Velocity*, PGV, Housner *Spectral Intensity*, *Arias Intensity*, fault mechanism and effective time duration of the record. The *significant method* is used for determining effective duration of earthquakes [20]. In addition, *Arias Intensity* and *Housner Intensity* measures are considered as the attributes related to the energy of the records. All the accelerograms are scaled to PGA of 0.35g prior to further spectral analysis.

In the present study, 100 earthquakes with magnitude of at least 5 Richter are employed; half of which being *Near-Field* (NF) records with epicentral distances not exceeding 20 kilometers. The other 50 earthquakes are labeled as *Far-Field* (FF) records to form the second record matrix. Every such *Earthquake Record Matrix*, ERM, is distinctly clustered by KM and ECBO-KM algorithms.

Table 1: Control parameters of the hybrid clustering framework

N	CMS	P_m	N_{KMS}	$N_{MaxIter}$
20	7	0.25	1000	1500

[DOI: 10.22068/ijoce.2025.15.1.621]

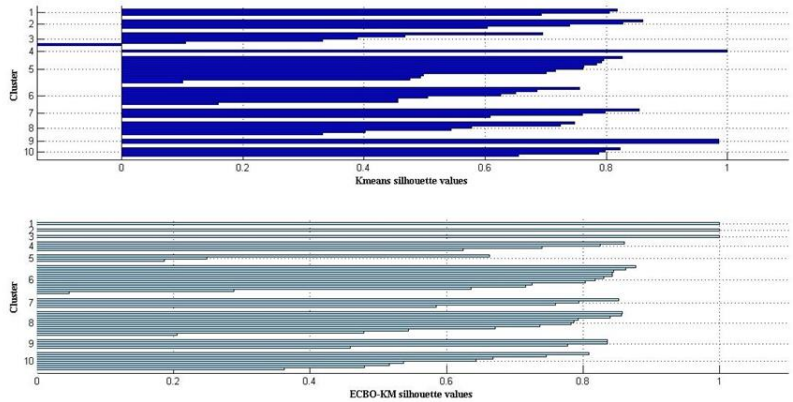


Figure 2: Silhouette plots for partitioning the NF earthquakes into 10 clusters

Clustering is performed for different cases of K=4 to K=11. ECBO-KM is run several times; for which Table 1 gives the successive control parameters. In each case, KM algorithm is run with the same population size as ECBO-KM to have a fair comparison [21,22].

The best results of KM and ECBO-KM are compared for different number of clusters in Table 2. For both NF and FF data groups, the best fitness and silhouette-sum by KM is generally less than ECBO-KM. The matter is highlighted for larger K values.

It is evident in the sample plots of Fig.2 and Fig.3 that KM has not been successful in avoiding negative silhouette values while ECBO-KM has resulted in considerably more desirable results. Note that, a proper clustering corresponds to higher silhouette values (closer to +1) and vice versa. Best achieved fitness results of Table 2 shows that such a non-deterministic search has been more successful than deterministic KM in overpassing local optima toward global optimum. In another word, the global search capability has been enhanced by hybridizing KM within ECBO.

Table 2: The best achieved fitness in clustering of earthquakes

K	Earthquake Field	KM	ECBO-KM
4	NF	0.6296	0.7457
	FF	0.5013	0.5268
5	NF	0.6546	0.6923
	FF	0.4968	0.5816
6	NF	0.6651	0.6549
	FF	0.5419	0.5449
7	NF	0.6331	0.6902
	FF	0.5772	0.6018
8	NF	0.6513	0.6908
	FF	0.5821	0.6245
9	NF	0.6133	0.6662
	FF	0.5932	0.6208
10	NF	0.6315	0.6898
	FF	0.5816	0.6244
11	NF	0.6154	0.6941
	FF	0.6149	0.6159

[DOI: 10.22068/ijoce.2025.15.1.621]

Table 3: Properties of the 3-bay 3-story moment frame (3ST model)

Story	Beam Sections	Column Sections
1 to 3	HEB240	IPE300

Table 4: Properties of the 3-bay 9-story moment frame (9ST model)

Story	Beam Sections	Column Sections
1	HEB340	IPE360
2 to 5	HEB340	IPE400
6, 7	HEB320	IPE360
8, 9	HEB300	IPE330

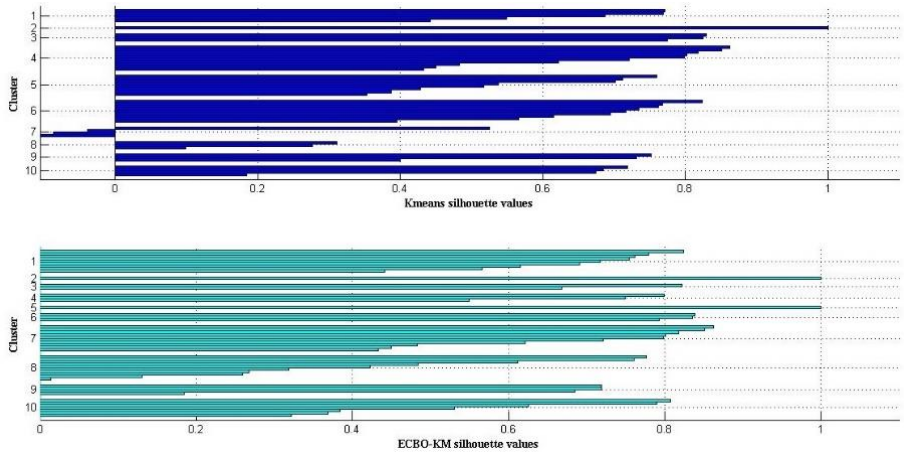


Figure 3: Silhouette plots for partitioning the FF earthquakes into 10 clusters

Table 5: Properties of the 3-bay 15-story moment frame (15ST model)

Story	Beam Sections	Column Sections
1	HEB500	IPE300
2, 3	HEB500	IPE400
4, 5	HEB500	IPE450
6, 7	HEB450	IPE400
8 to 12	HEB400	IPE400
13, 14	HEB400	IPE360
15	HEB400	IPE330

4.2 Optimal clustering of earthquakes due to structural responses

Another way of clustering earthquakes is to distinguish their similarities or differences after they are filtered by structural models; that is clustering of consequent responses. In the present study, NF and FF group of acceleration records are first normalized to 0.35g and then

[DOI: 10.22068/ijoce.2025.15.1.621]

distinctly applied to a number of literature examples. These are moment frames designed for PGA=0.35g and soil class B at the construction site according to EC8 [23]. The structural profiles of the 3 story, 9 story and 15 story frames are reported in Tables 3 to 5, respectively. Each of the frames has 3 bays. They are here-in-after denoted by 3ST, 9ST and 15ST models, respectively. Material property is fixed to the constructional steel with the yield stress of 235MPa. Identical uniform load of 27.5kN/m is applied on every frame beams.

Columns are rigidly connected to beams. Taking into account that the structure is inelastic, the damping matrix changes with the stiffness variation during the time history analysis. In this regard, a Rayleigh damping of 5% is utilized for the first and the last modes of vibration. Dynamic nonlinear time-history analyses are applied by distributed plasticity element in OpenSees software using the constant acceleration method for the numerical integration [24].

For each case of clustering, a data matrix called *Structural Response Matrix*, SRM, is provided that reflects seismic characteristics of the treated frames. It includes the following attributes:

I) Weighted average of the maximal story drift, story displacement and column stress responses over time-points of the non-linear analysis when story masses are taken the averaging weights:

$$D_{\mathrm{I}} = \frac{\sum_{i=1}^{N} M_i \times DS_i}{\sum_{i=1}^{N} M_i} \tag{14}$$

$$Dr_{\mathrm{I}} = \frac{\sum_{i=1}^{N} M_i \times DF_i}{\sum_{i=1}^{N} M_i} \tag{15}$$

$$R_{\mathrm{I}} = \frac{\sum_{i=1}^{N} M_i \times Cr_i}{\sum_{i=1}^{N} M_i} \tag{16}$$

N indicates the number of stories while M_i and Cr_i denote the mass and the maximum column stress ratios in the i^{th} story, respectively. DS_i and DF_i are the corresponding maximal displacement and drift ratios.

II) Simple mean of the maximal story drift, displacement and column stresses during the nonlinear analysis:

$$D_{\mathrm{II}} = \frac{\sum_{i=1}^{N} DS_i}{N} \tag{17}$$

$$Dr_{\mathrm{II}} = \frac{\sum_{i=1}^{N} DF_i}{N} \tag{18}$$

$$R_{\mathrm{II}} = \frac{\sum_{i=1}^{N} Cr_i}{N} \tag{19}$$

[DOI: 10.22068/ijoce.2025.15.1.621]

III) Maximum structural responses among all stories over time, derived by time-history analysis:

$$D_{III} = \max_t Roof\ \ Displacement \tag{20}$$

$$Dr_{\mathrm{III}} = \max_t Dr \tag{21}$$

$$R_{\mathrm{III}} = \max_t Cr_i \tag{22}$$

Maximum resulted base shear during nonlinear dynamic analysis is considered as the last attribute in the SRM. Therefore, in each case of near-field or far-field earthquakes, for each frame model, such a matrix is constructed by $Ne = 50$ rows and 10 columns.

After performing non-linear dynamic analyses, the corresponding SRM's are generated for further clustering.

Table 6 gives the results of all 48 cases of clustering by KM and ECBO-KM. The results of optimal clustering on several SRM's, confirm superior performance of ECBO-KM over KM algorithm.

5. COHERENCY MEASUREMENT

Time-history records of strong ground motion have the role of input signals to the structural system and carry special information of the corresponding earthquake. Such seismic signals are filtered by the structure to derive consequent responses. Analogous to light rays in optics, the earthquake signals may be coherent or incoherent to each other. Clustering provides a mathematical tool to derive coherency between such seismic signals, before or after being filtered by the structural model.

Once earthquake records and consequent structural responses are clustered, some further issues arise to be investigated. For example: how much coherency of clusters is preserved before and after structural analysis and which degree of similarity remains between the seismic clusters filtered by different structures under similar earthquake excitations. The matter necessitates deriving a metric for record-to-response and response-to-response coherency measurements. The present work attempts to derive such a metric, via the following steps:

- Determine two case of clustering A and B for which the coherency is to be measured. For example, B may indicate the clustering of FF records where A can be clustering of the corresponding seismic responses in the specific model. Run the proposed hybrid framework to obtain K clusters of each case as:

$$A = \left\{C_1^A, ..., C_K^A\right\}, B = \left\{C_1^B, ..., C_K^B\right\} \tag{23}$$

- Form the matrix T so that every its component T_{ij} is defined as the number of identical entities (earthquakes) between C_i^A and C_j^B.
- Re-arrange T so that for every its column the maximum-value component falls on the diagonal.

[DOI: 10.22068/ijoce.2025.15.1.621]

Table 6: The best fitness in clustering of the structural responses under NF and FF seismic excitations

Number of clusters	Earthquake's field	Structural Model	KM	ECBO-KM
			Fitness	Fitness
4	NF	3ST	0.7635	0.7649
		9ST	0.7307	0.7965
		15ST	0.7310	0.7482
	FF	3ST	0.7854	0.803
		9ST	0.7967	0.7967
		15ST	0.7197	0.7526
5	NF	3ST	0.7028	0.8043
		9ST	0.7653	0.7628
		15ST	0.7375	0.7375
	FF	3ST	0.7673	0.7960
		9ST	0.7714	0.7753
		15ST	0.7332	0.7332
6	NF	3ST	0.7052	0.7688
		9ST	0.7389	0.7606
		15ST	0.7691	0.7691
	FF	3ST	0.7602	0.8052
		9ST	0.7265	0.7602
		15ST	0.6761	0.7507
7	NF	3ST	0.6761	0.7928
		9ST	0.7476	0.7596
		15ST	0.7746	0.7458
	FF	3ST	0.7705	0.7945
		9ST	0.7361	0.7454
		15ST	0.6984	0.7406
8	NF	3ST	0.6879	0.7785
		9ST	0.7048	0.7727
		15ST	0.7232	0.7689
	FF	3ST	0.7803	0.8251
		9ST	0.7441	0.7682
		15ST	0.6687	0.7707
9	NF	3ST	0.6745	0.692
		9ST	0.729	0.7785
		15ST	0.7595	0.7788
	FF	3ST	0.8003	0.8264
		9ST	0.7306	0.7474
		15ST	0.7377	0.7606
10	NF	3ST	0.6709	0.7072
		9ST	0.6874	0.7663
		15ST	0.7547	0.7591
	FF	3ST	0.7718	0.8362
		9ST	0.7567	0.7539
		15ST	0.765	0.7773
11	NF	3ST	0.6990	0.7117
		9ST	0.6702	0.7416
		15ST	0.8010	0.819
	FF	3ST	0.8043	0.8413
		9ST	0.7649	0.7538
		15ST	0.7734	0.7843

[DOI: 10.22068/ijoce.2025.15.1.621]

- If T_{ij} is not the greatest value among its row and column, it is inevitably be substituted with zero. Eliminate off-diagonal components of such a *Coherency Matrix*, CM.
- Calculate the *Coherency Index* between A and B as:

$$CI = \frac{\sum_{i=1}^{K} CM_{ii}}{N_e} \tag{24}$$

Tables 7 and 8 report structural responses-to-record CI for near-field and far-field ERM's, respectively. These constitute 48 independent comparison cases; for which the corresponding CI has fallen below 50%.

Among all 48 cases, one is sampled for illustrative purposes; that is the case of K=7 when A denotes seismic-response clustering of the 15ST model and B corresponds to the NF set of excitations. The resulted pattern of CM is like Fig.4 while the corresponding clusters with similar earthquakes, are reported in Table 9. It is evident that there are some C_i^A and C_j^B that do not have any identical earthquake; i.e. they are completely incoherent to each other.

Table 7: Response-to-record CI by ECBO-KM clustering between structural responses and NF records

		A		
K		3ST	9ST	15ST
4		40%	50%	44%
5		48%	38%	38%
6		44%	40%	38%
7	B	36%	38%	46%
8		36%	40%	42%
9		34%	40%	36%
10		34%	34%	38%
11		32%	40%	40%

Table 8: Response-to-record CI by ECBO-KM clustering between structural responses and FF records

		A		
K		3ST	9ST	15ST
4		50%	40%	40%
5		36%	30%	38%
6		36%	36%	46%
7	B	32%	38%	42%
8		34%	42%	44%
9		38%	40%	42%
10		36%	42%	42%
11		28%	38%	44%

[DOI: 10.22068/ijoce.2025.15.1.621]

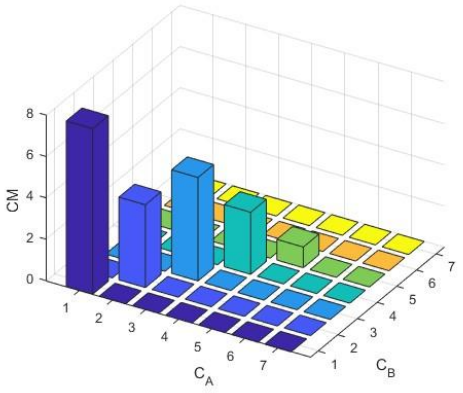


Figure 4: Sample pattern of CM

It can be realized that CI not only depends on the structural model and the field of earthquake records but also is affected by the number of clusters. For example, the 3ST model has its maximal response-to-record CI at K=5 in NF and K=4 in FF excitations. However, it is not the case for taller frame models; e.g. 15ST has its greatest CI at K of 6 and 7 under NF and FF seismic excitations, respectively.

It can be noticed in Fig.5 that the spectral responses of different clusters are diffracted from each other. In another word, the entities in each cluster are coherent to each other but incoherent to the other entities in view of spectral responses. Such a phenomenon is highlighted for medium-to-high periods; especially in the velocity spectrum which is related to the input energy of the record. The matter confirms that records in different clusters appear in incoherent spectral levels after being filtered via non-linear structural analyses. Moreover, such a diffraction agrees with the proposed A-to-B coherency pattern. In practical point of view, such a coherency picture acts as a guide for selecting one of the coherent earthquake records as representative of that cluster when providing the input set of records. It is a need for further seismic analysis and design.

Fig. 6 reveals comparison of CI metric for 24 structural response-to-record cases under near-field records. It can be noticed that response-to-response CI values have generally been higher than response-to-record ones for each structural model. For the case of far-filed (FF) excitation, CI values are given in Fig.7. Comparison of results in Fig.6 with those in Fig.7 declares that response-to-response CI has generally experienced greater variation under FF excitation than NF for a fixed K. In the case of 9 clusters under NF excitation; e.g. CI has been obtained 44%, 34% and 38% for 3ST-to-9ST, 3ST-to-15ST and 9ST-to-15ST pairs, respectively. For FF excitation; however, the corresponding CI values have been changed to 32%, 28% and 60%. The maximum CI of 62% has occurred between response clusters of the 3ST and 9ST models under FF excitation when K is 4. Range of CI variation is 34~40% for NF excitation; that is considerably lower than 24~62% for FF set of records.

It is also declared that of 3ST-to-15ST is generally lower than the other two cases. However, for FF excitation and K=5 there is an exception with minor difference in CI of 3ST-to-15ST with respect to 9ST-to-15ST. It is also noticeable that the number of K cases that CI of the 9ST-to-15ST considerably overrides the other two, has increased under the FF excitation with respect to NF.

[DOI: 10.22068/ijoce.2025.15.1.621]

Table 9: Sample A-to-B similar clustered earthquakes in deriving CM for K=7, 15ST frame

Cluster ID	Earthquake ID	A		B	
		Silhouette	Average	Silhouette	Average
1	San Fernando-Lake Hughes #12	0.2586	0.613	0.2586	0.667
	San Francisco-Golden Gate Park	1		1	
	Norcia, Italy- Spoleto	0.8501		0.8501	
	HelenaMontana-01-Carroll College	0.4537		0.4537	
	Chi-ChiTaiwan-03-TCU073	0.5478		0.5478	
	Dursunbey, Turkey-Dursunbey	0.5782		0.5782	
	Mammoth Lakes01-Mammoth Lakes	0.6233		0.6233	
	Tottori Japan-OKY004	0.5906		0.5906	
2	Imperial Valley-06-El Centro Array #1	0.8657	0.694	0.8657	0.656
	Imperial Valley-04-El Centro Array #9	0.8848		0.8848	
	ManaguaNicaragua-02-ManaguaESSO	0.3813		0.3813	
	ManaguaNicaragua-01-ManaguaESSO	0.307		0.307	
	Hollister-01-Hollister City Hall	0.87		0.87	
	Imperial Valley-06-Chihuahua	0.8582		0.8582	
3	DarfieldNew Zealand-DFHS)	0.7045	0.751	0.7045	0.817
	Umbria Marche Italy-Bevagna	0.8065		0.8065	
	Mammoth Lakes-02-Convict Creek	0.8606		0.8606	
	N. Palm Springs-Cabazon	0.7768		0.7768	
	Chuetsu-oki Japan-Ojiya City	0.506		0.506	
	Chuetsu-oki Japan_NIG019	0.8517		0.8517	
4	San Simeon CA-San Antonio Dam - Toe	0.9046	0.860	0.9046	0.925
	Loma Prieta-Anderson Dam (L Abut)	0.8453		0.8453	
	Chuetsu-oki Japan-Tani Kozima Nagaoka	0.8312		0.8312	
5	Lytle Creek-Devil's Canyon	0.472	0.472	0.939	0.939
6	-	-		-	
7	-	-		-	

[DOI: 10.22068/ijoce.2025.15.1.621]

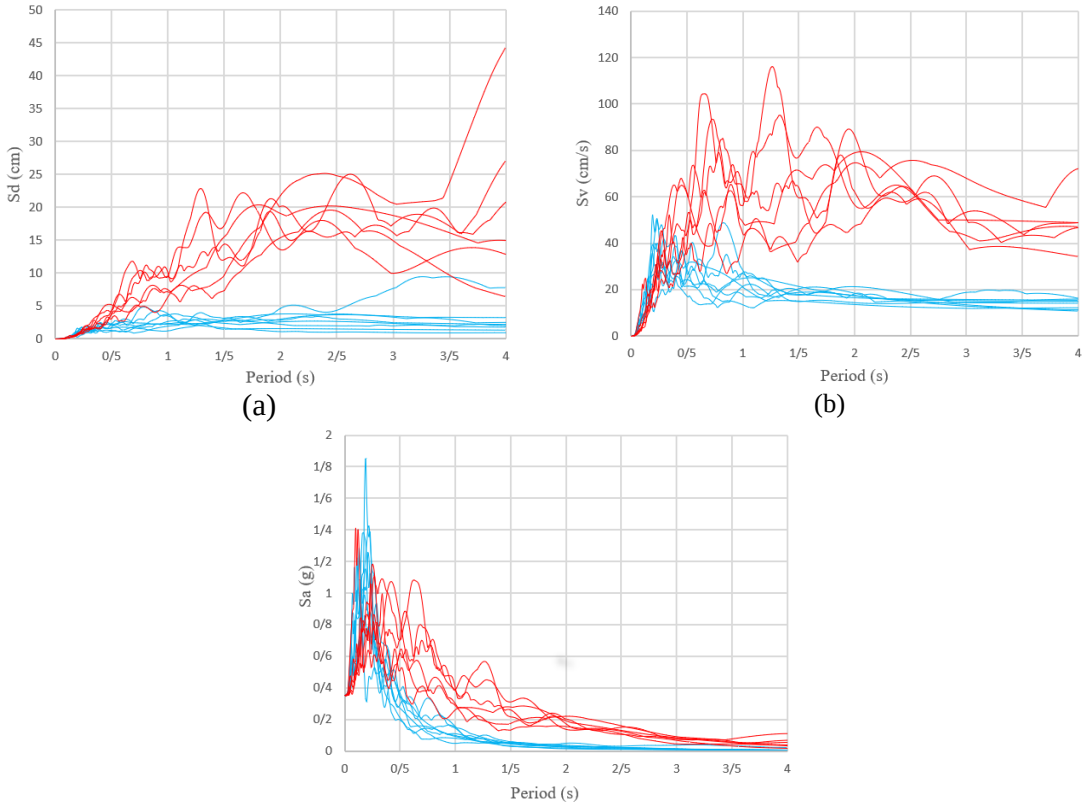


Figure 5: Spectral responses of 15ST model in the 1st and 2nd clusters including (a) Sd , (b) Sv and (c) Sa. Incoherent entities are distinguished in different inks.

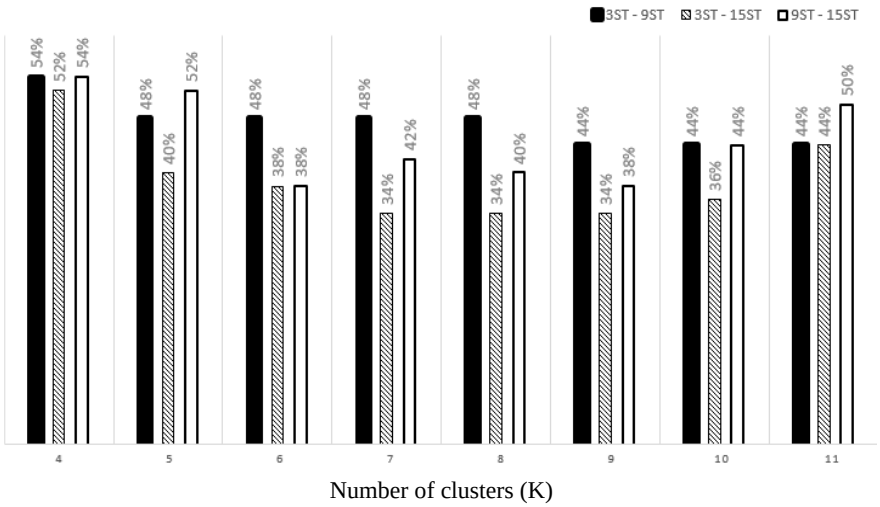


Figure 6: Response-to-response CI results for NF excitations after optimal clustering

[DOI: 10.22068/ijoce.2025.15.1.621]

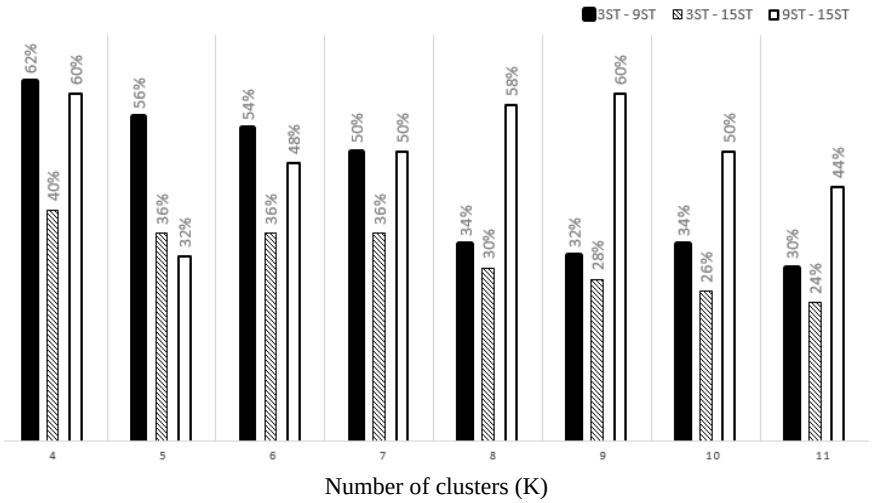


Figure 7: Response-to-response CI results for FF excitations after optimal clustering

6. CONCLUSION

The present work proposed a novel metric to take into account both the structural response and input excitation coherency. Two types of optimal clustering were studied: the first concerns earthquake attributes which address geotechnical and seismic data, while the second focuses on clustering of the resulting structural responses.

Optimization was employed to provide uniqueness in clustering for better conclusion on coherency measurements. Since premature convergence to local optima was detected in application of KM algorithm, it was hybridized as a local search with a variant of CBO to maintain better search, within an optimal clustering framework. Numerical results show that the proposed hybrid method can provide proper search refinement toward global optima in both types of clustering on the earthquake records or on the seismic-responses, under either near-field or far-field excitations.

CI for record-to-response cases, was generally obtained lower than for the case of response-to-response coherency. In addition, structural height was found to have considerable effect on the latter case. The more difference in the height of the frame model, the more variation in CI was observed.

As another issue, CI variation between far-filed and near-field earthquake records were studied. It was found that the second case generally causes more CI variation between various models and especially different K values. According to the present study, far-field records can lead to higher response-to-response coherency; compared with the corresponding cases for near-field excitation. Greater sensitivity of CI to the structural height and the number of clusters was observed for the far-field case. The proposed procedure was confirmed by observing incoherency of different clusters in spectral seismic responses. Such a difference was more in the velocity response spectra than the acceleration spectra. As a future scope of work, the proposed index can be evaluated on more structural systems in various types and richer database of seismic excitation.

[DOI: 10.22068/ijoce.2025.15.1.621]

REFERENCES

1. Larose D. T. *Discovering Knowledge in Data: An Introduction to Data Mining.* Wiley Blackwell; 2005.
2. Zaki M. J., Meira M. J. *Data Mining and Analysis: Fundamental Concepts and Algorithms.* 2013.
3. Eick C. F., Zeidat N., Vilalta R. Using representative-based clustering for nearest neighbor dataset editing. *Proc 4th IEEE Int Conf Data Min (ICDM).* 2004; 375–8.
4. Sarkar S., Roy A., Purkayastha B. S. Application of Particle Swarm Optimization in Data Clustering: A Survey. *Int J Comput Appl.* 2013; **65**:975–8887.
5. Naeim F., Alimoradi A., Pezeshk S. Selection and scaling of ground motion time histories for structural design using Genetic Algorithms. *Earthq Spectra.* 2004; **20**:413–26.
6. Shahrouzi M., Mohammadi A. Optimal ground motion scaling using enhanced swarm intelligence for sizing design of steel frames. *Int J Optim Civ Eng.* 2014; **4**:293–308.
7. Fahjan Y. M. Selection and scaling of real earthquake accelerograms to fit the Turkish design spectra. *Tek Dergi Tech J Turkish Chamb Civ Eng.* 2008; **19**:1231–50.
8. Hartigan J. A., Wong M. A. A K-Means Clustering Algorithm. *J R Stat Soc Ser C.* 1975; **28**:100–8.
9. Zhang J., Lin X. An Adaptive Ant Colony Clustering Algorithm and Application in the TSP. *Proc 2022 Int Conf Comput Network Electron Autom (ICCNEA).* 2022; 82–5.
10. Rana S., Jasola S., Kumar R. A review on particle swarm optimization algorithms and their applications to data clustering. *Artif Intell Rev.* 2011; **35**:211–22.
11. Köhler A., Ohrnberger M., Scherbaum F. Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps. *Geophys J Int.* 2010; **182**:1619–30.
12. Kaveh A., Fahimi-Farzam M., Kalateh-Ahani M. Performance-based multi-objective optimal design of steel frame structures: Nonlinear dynamic procedure. *Sci Iran.* 2015; **22**:372–87.
13. Ahmadi M., Attari N. K. A., Shahrouzi M. Structural seismic response mitigation using optimized vibro-impact nonlinear energy sinks. *J Earthq Eng.* 2015; **19**:193–219.
14. Kaveh A., Mahdavi V. R. Colliding bodies optimization: A novel meta-heuristic method. *Comput Struct.* 2014; **139**:18–27.
15. Kaveh A., Mahdavi V. R. *Colliding Bodies Optimization: Extensions and Applications.* 2015; 1–284.
16. Kaveh A., Ardebili S. R. A Comparative Study of the Optimum Tuned Mass Damper for High-rise Structures Considering Soil-structure Interaction. *Period Polytech Civ Eng.* 2021; **65**:1036–49.
17. Kaveh A. *Advances in Metaheuristic Algorithms for Optimal Design of Structures.* 3rd ed. Cham: Springer International Publishing; 2021.

[DOI: 10.22068/ijoce.2025.15.1.621]

18. Shahrouzi M., Rashidimoghadam M. Ground motion clustering by a hybrid K-means and colliding bodies optimization. *Int J Optim Civ Eng.* 2016; **6**:567–78.
19. Kaveh A., Ilchi Ghazaan M. Enhanced colliding bodies optimization for design problems with continuous and discrete variables. *Adv Eng Softw.* 2014; **77**:66–75.
20. Elnashai A. S., Di Sarno L. *Fundamentals of Earthquake Engineering.* John Wiley & Sons Inc.; 2008.
21. Shahrouzi M., Kaveh A. An efficient derivative-free optimization algorithm inspired by avian life-saving maneuvers. *J Comput Sci.* 2022; **57**:101483.
22. Taghavi A. M., Shahrouzi M. Optimal design of spatial structures by a novel meta-heuristic algorithm: Sound energy optimizer. *Structures.* 2024; **70**:107570.
23. Karavasilis T. L., Bazeos N., Beskos D. E. Behavior factor for performance-based seismic design of plane steel moment resisting frames. *J Earthq Eng.* 2007; **11**:531–59.
24. Mazzoni S., McKenna F., Scott M. H., Fenves G. L. *OpenSees Command Language Manual.* Pacific Earthq Eng Res Cent; 2007; 451.

[DOI: 10.22068/ijoce.2025.15.1.621]